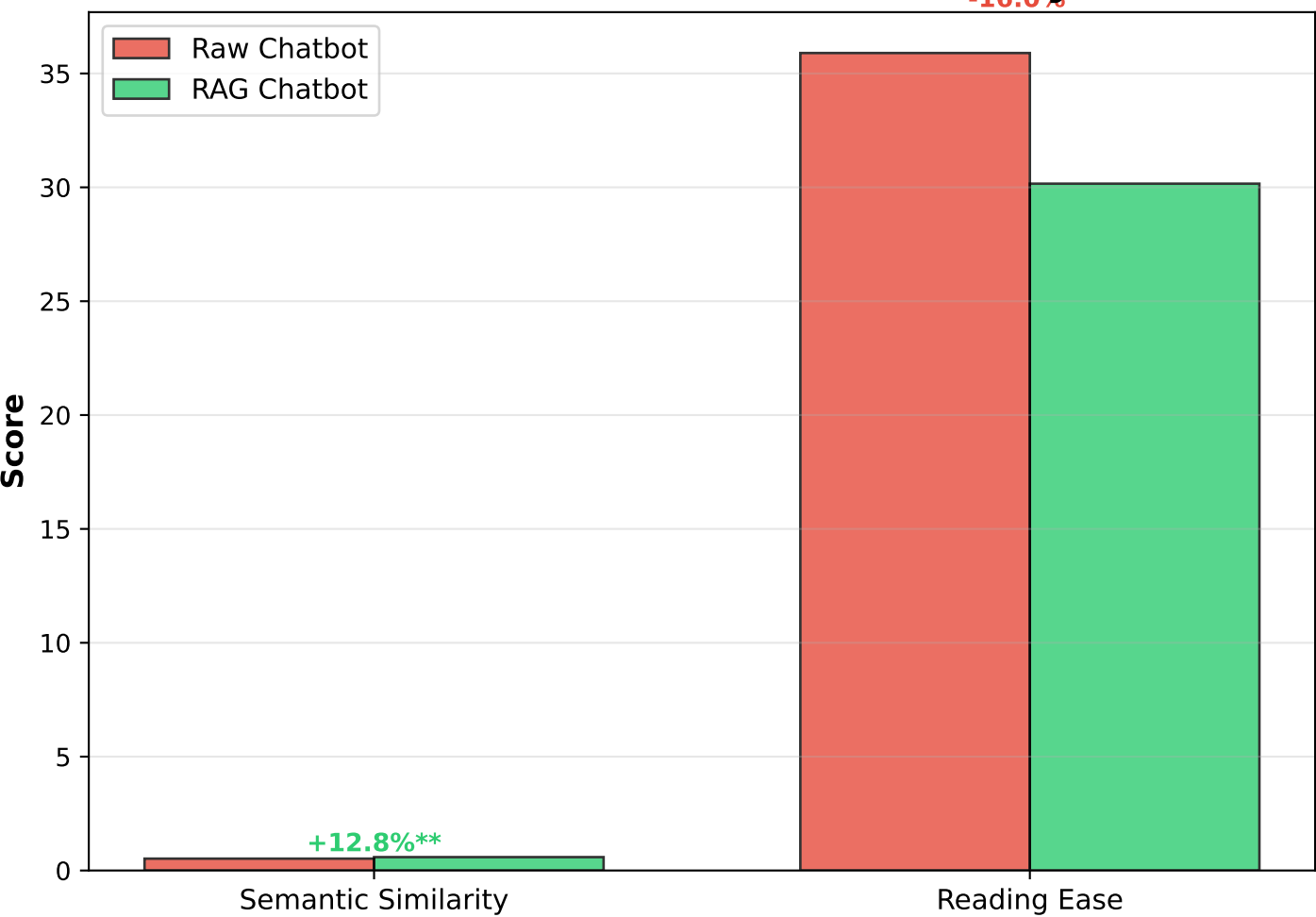
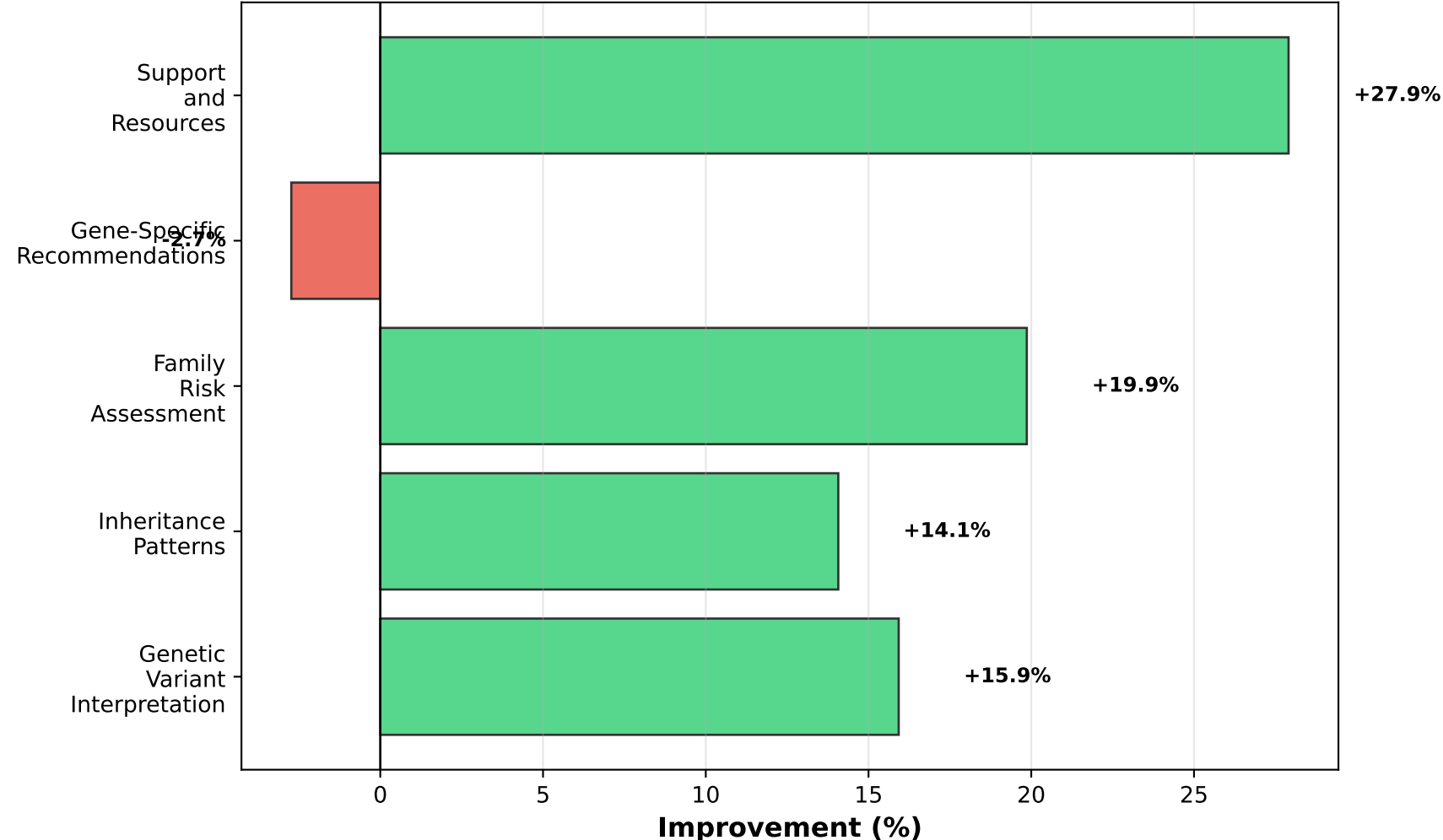


CGT Chatbot Analysis: RAG vs Raw Performance

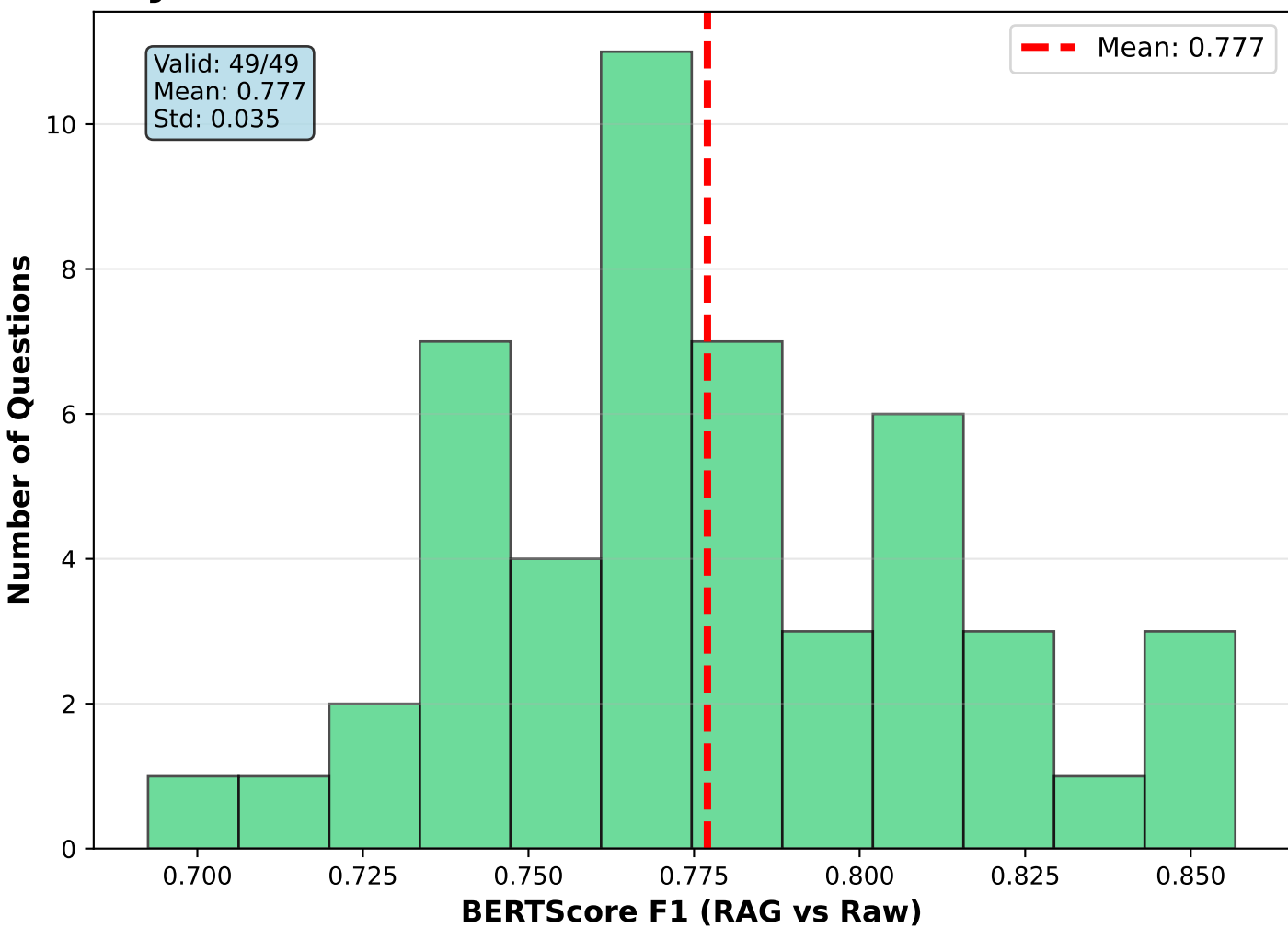
Overall Performance Comparison



Semantic Similarity Improvement by Category



BERTScore F1 Distribution



Performance Summary

Metric	Raw	RAG	Change	Significant
Semantic Similarity	0.521	0.588	+12.8%	✓
Reading Ease	35.9	30.2	-16.0%	✓
BERTScore F1	N/A	0.777	N/A	N/A

Key Findings:

- 49 questions across 5 categories
- BERTScore F1: 0.777 ± 0.035
- Best category: Support and Resources